



Identification des sources de variabilité des niveaux de polluants par analyse en composantes indépendantes et par estimation de mélange des données de l'air intérieur

Rachid Ouaret, Anda Ionescu, Olivier Ramalho, Viorel Petrehus, Yves Candau

► To cite this version:

Rachid Ouaret, Anda Ionescu, Olivier Ramalho, Viorel Petrehus, Yves Candau. Identification des sources de variabilité des niveaux de polluants par analyse en composantes indépendantes et par estimation de mélange des données de l'air intérieur. Journées Interdisciplinaires de la Qualité de l'Air, Feb 2014, Villeneuve d'Ascq, France. 18 p. hal-01064376

HAL Id: hal-01064376

<https://hal.science/hal-01064376>

Submitted on 16 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identification des sources de variabilité des niveaux de polluants par analyse en composantes indépendantes et par estimation de mélange des données de l'air intérieur

R.OUARET^{1,2}, A. IONESCU², O.RAMALHO¹, V. PETREHUS³, Y. CANDAU²

rachid.ouaret@cstb.fr

¹ Centre Scientifique et Technique du Bâtiment

² Centre d'Études et de Recherche en Thermique, Environnement et Systèmes (CERTES EA 3481), Université Paris-Est Créteil

³ Université Technique des Constructions de Bucarest, Département de Mathématiques, Roumanie

Mots-clés: Formaldéhyde, Particules, Analyse en composantes indépendantes, modèle de mélange.

1 Introduction

L'identification des sources de pollution dans les environnements intérieurs et la quantification de leur contribution nécessite de développer des connaissances sur la variabilité temporelle de la concentration en polluants dans l'air. Depuis plusieurs années, les méthodes de séparation des sources "en aveugle" ont été essentiellement développées et appliquées dans le domaine de l'air extérieur ou du traitement du signal. Étant donné que les concentrations de certains polluants sont en général plus élevées à l'intérieur qu'à l'extérieur, notamment pour le formaldéhyde, et que les environnements intérieurs représentent 80% de l'exposition totale aux polluants de l'air, il est devenu nécessaire de mieux comprendre les fluctuations par l'évaluation des contributions des différentes sources pour un polluant donné.

Dans ce travail, une étude sur la caractérisation des particules émises par différentes sources potentielles dans une maison expérimentale a été entreprise. Pour cela des mesures de formaldéhyde et de particules à l'intérieur et à l'extérieur d'une maison instrumentée (MARIA, située sur le site du CSTB à Champs sur Marne) ont été effectuées avec un pas de temps d'une minute, du 19/04/2010 à 09:32 au 23/04/2010 à 23:59. Les concentrations des particules dans l'air intérieur sont enregistrées par un compteur optique des particules sur 15 gammes de taille entre 0.3 et 20 μm . Les particules dans l'air extérieur sont enregistrées sur 8 gammes de taille entre 0.3 et 2.5 μm . Des mesures de concentration formaldéhyde (en ppb) ont été réalisées par un analyseur en temps réel du formaldéhyde dans l'air (Aerolaser GmbH). Au total, on dispose de 24 variables sur une période de 4 jours.

L'objectif général de cette étude est de d'utiliser plusieurs jeux de données pour établir des modèles de fluctuations des sources de polluants par deux approches, une appliquée aux fluctuations des concentrations de formaldéhyde et l'autre aux fluctuations en nombre de particules par unité de volume.

Ce travail présente une modélisation statistique de l'identification des sources de la variabilité temporelle des concentrations de ces polluants par deux méthodes : analyse en composantes indépendantes (ACI) et modèle de mélange gaussien. L'approche par l'ACI relève du problème de la séparation aveugle des sources, qui consiste à retrouver n sources inconnues observées au travers de m mélanges de ces sources.

Les 35 variables (polluants et paramètres climatiques intérieurs et extérieurs) constituent un ensemble des séries temporelles que nous avons prétraité en développant une méthodologie d'imputation des valeurs manquantes, notamment pour le formaldéhyde.

L'objectif de notre travail, dont cet article n'est qu'une première étape, est de montrer en quoi certains outils statistiques les plus récents peuvent apporter des éclairages nouveaux sur la compréhension des fluctuations des sources particulières dans un environnement intérieur.

2 Matériels et méthodes

2.1 Description du bureau instrumenté

La maison MARIA (**M**aison **A**utomatisée pour la **R**echerche et **I**nnovation) du CSTB (**C**entre **S**cientifique et **T**echnique du **B**âtiment) situé à Champs sur Marne, en Ile de France, est une maison individuelle de type F5 (Figure 2-1). Elle possède cinq pièces principales, un séjour et quatre chambres ainsi que deux toilettes, une salle de douche et une salle de bain, réparties sur deux niveaux.



Figure 2-1 Vue de la façade sud de la maison expérimentale MARIA

Elle a une surface habitable d'environ 142 m^2 et un volume total (sans la toiture) d'environ 656 m^3 . La maison MARIA est quasi-automatisée : les ouvertures sont automatisées et la perméabilité est contrôlée avec une ventilation modulable ; de plus, les appareils ménagers peuvent être gérés depuis un poste de commande [1]. Le système de ventilation a été fixé en mode simple flux avec des bouches d'extraction disposées dans les pièces (douche, salle de bain, toilettes et cuisine).

Par ailleurs, une station météorologique placée sur le toit permet de mesurer en continu les paramètres environnementaux : la température et l'humidité extérieure, la vitesse et la direction du vent, l'ensoleillement, la pluviométrie et la pression atmosphérique. Cette instrumentation est conçue de façon à compléter la mesure de la qualité de l'air intérieur de la maison.

2.2 Mesures et statistiques des polluants

2.2.1 Mesures en continu

Dans le cadre de l'étude, la campagne de mesure consiste en la réalisation d'une base de données de plusieurs polluants avec un pas de temps d'une minute. La

base de données originale possède des valeurs manquantes ; c'est pourquoi nous privilégions une plage de mesures complète afin de ne pas entacher l'information d'erreurs après imputation des données. Après prétraitement, le polluant formaldéhyde présente 21 % de valeurs manquantes qui ont été remplacées par interpolation.

La base de données et les appareils de mesures utilisés sont présentés dans le Tableau 2-1.

Environnement	Type	Paramètres mesurés	Instrument de mesure
Intérieur	Polluants	HCHO	AL 4021 (Aerolaser GmBh) permet d'enregistrer la concentration de formaldéhyde dans l'air au niveau du ppb (limite de détection de 0,05 ppb)
		15 gammes de particules (0.35 à >20µm)	Dust Monitor 1.108, (Grimm).Compteur optique des particules
		CO ₂	Sonde q-track
	Paramètres environnementaux	Température	
		l'humidité relative	
Extérieur	Polluants	8 gammes Particules (0.35 à 2.5µm)	Dust Monitor 1.108, (Grimm).Compteur optique des particules
		CO ₂	Station météo
	Paramètres environnementaux	Température	
		Humidité relative	

Tableau 2-1 Tableau récapitulatif des paramètres mesurés et de l'instrumentation.

Afin de représenter au mieux les expositions des occupants, les mesures de formaldéhyde et des particules ont été effectuées à la hauteur des voies respiratoires et sur la même période pour tous les appareils.

Au total, 26 polluants intérieurs et extérieurs ont été mesurés sur une période de 4 jours (du 19/04/10 à 09:32 au 23/04/10 à 23:59). Pour les gammes de particules entre 0.3-2.5 µm (8 classes) correspondant aux mesures de l'air extérieur, le ratio I/E correspondant au rapport entre la concentration intérieure et celle de l'extérieur a été calculé.

Outre la mesure des concentrations en polluants cibles (particules et HCHO), les paramètres climatiques intérieurs comme extérieurs ont été enregistrés durant la même période afin de garantir la représentativité physique de l'environnement étudié. Au total, huit paramètres climatiques intérieurs et extérieurs ont été mesurés : température (intérieure et extérieure), humidité relative (intérieure et extérieure), vitesse et direction du vent, pression atmosphérique et ensoleillement. Ces paramètres sont enregistrés de manière continue par l'application qui pilote la maison expérimentale et en intégrant les mesures de la station météorologique pour les informations extérieures.

2.2.2 Analyse statistique des variables

Les résultats de mesures sont présentés d'abord séparément par polluant et par paramètre climatique et ensuite simultanément par des méthodes d'analyse multivariée.

2.2.3 Le formaldéhyde

Les résultats de la statistique descriptive montrent que l'écart-type du formaldéhyde est de 2.6 ppb et la moyenne est de 14.31 ppb. Notons que l'erreur type de la moyenne (définie par le rapport entre l'écart type et la racine carrée de la taille de l'échantillon) est de 0.03 ppb est inférieure à la limite de détection de l'Aerolaser (l'analyseur du formaldéhyde) qui est de 0.05 ppb. L'intervalle de variation du HCHO est de 13.4 ppb avec un maximum de 22.1 ppb observé le 20/04/2010 à 18:44.

La distribution de probabilité du HCHO est légèrement aplatie (coefficient d'aplatissement, kurtosis égal à -0.5) et décalée à gauche de la médiane (coefficient d'asymétrie, skewness positif). La Figure 2-2 montre deux distributions de probabilité des niveaux du HCHO, l'une est exprimée en log et l'autre en valeurs brutes. On remarque, en dépit de la transformation en log, l'existence d'une structure polymodale pour les deux distributions ; cela laisse présager l'existence d'un mélange de distributions. Par conséquent, la distribution de ce jeu de données est prise à un mélange de composantes en nombre fini, qui sont modélisées par autant de fonctions de densité de probabilité. Nous verrons par la suite l'intérêt de cette modélisation dans le cas de l'estimation de la contribution de sources.

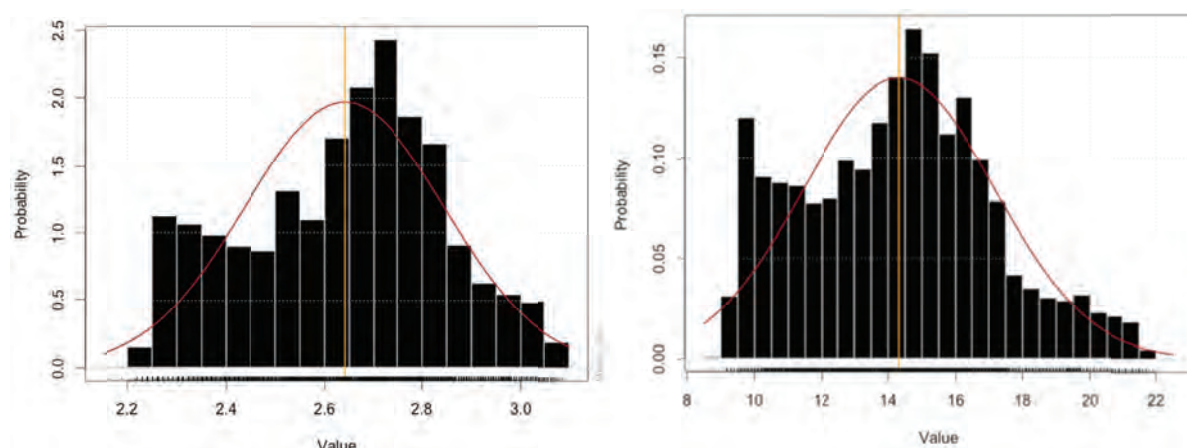


Figure 2-2 A droite : histogramme du HCHO des données brutes. A gauche : histogramme du log(HCHO).

Remarquons aussi que durant le premier jour (19/04/2010), les variations du formaldéhyde sont plus fluctuantes et présentent une forme différente que les variations observées durant les trois jours qui ont suivi (Figure 2-3).

Afin de pouvoir comparer la variabilité de l'humidité et celle du HCHO, on s'intéresse plutôt à l'humidité spécifique (absolue) qu'à celle relative. On rappelle que l'humidité spécifique représente la quantité de vapeur d'eau contenue dans l'air, et elle est exprimée en **g** de vapeur d'eau par **kg** d'air sec. La Figure 2-4 ci-dessous montre l'évolution temporelle de la concentration en formaldéhyde (en log) et la concentration de l'humidité spécifique intérieure. On remarque que la dynamique de la concentration du HCHO est similaire à celle d'humidité spécifique.

Les profils temporels des deux paramètres (Figure 2-3 ; Figure 2-4) exhibent une périodicité journalière, avec un minimum entre 4 et 7 heures du matin et un maximum de 14 à 19 heures pour HCHO et entre 12 et 14 heures pour l'humidité spécifique.

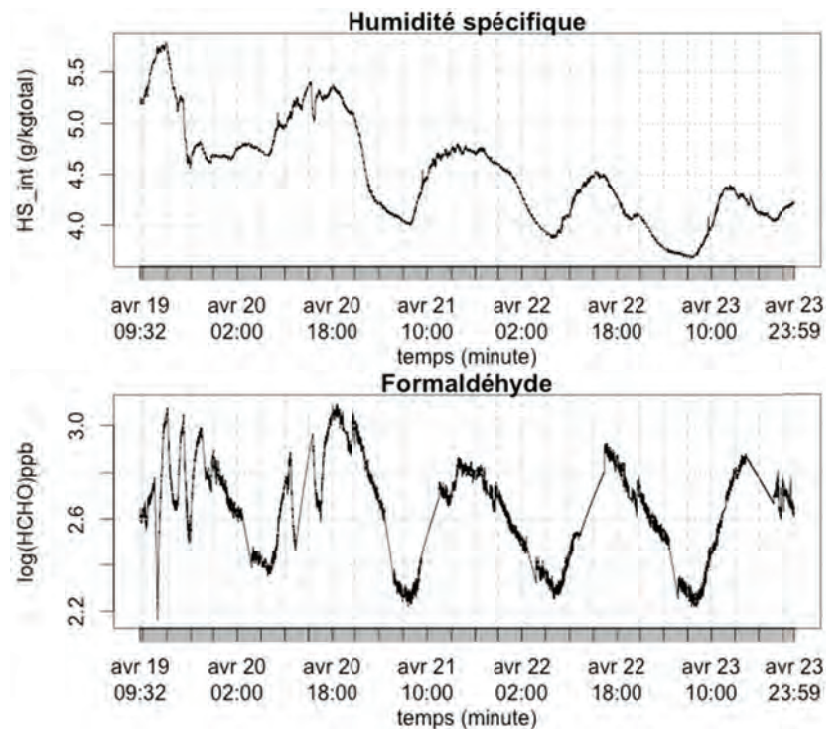


Figure 2-3 Évolution temporelle de séries HCHO (en log) et de l'humidité spécifique en g/kg total durant 4 jours.

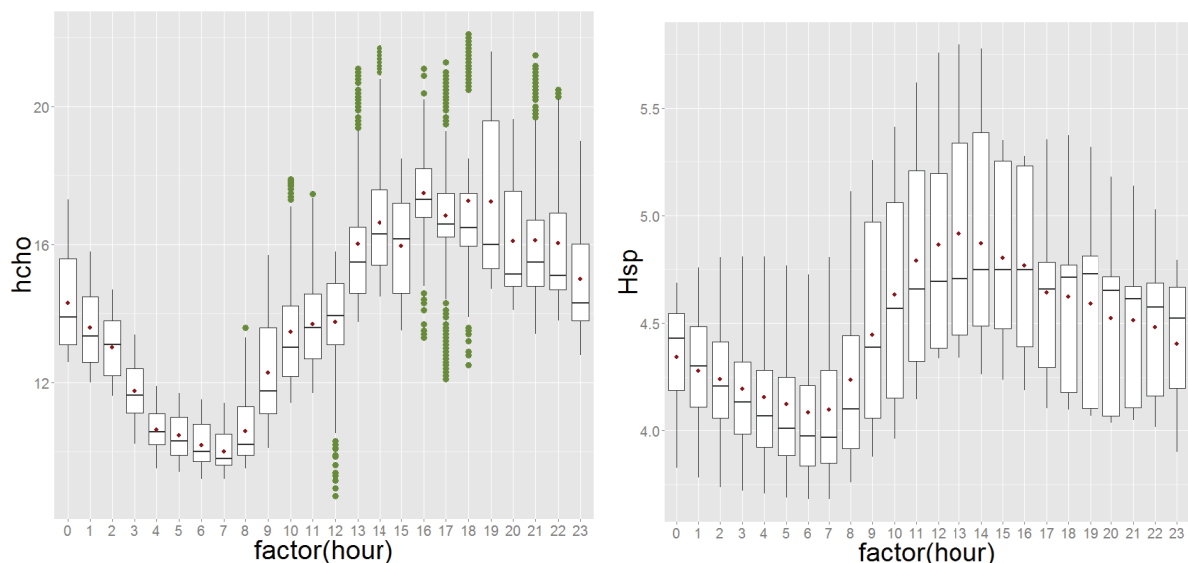


Figure 2-4 Profil diurne des concentrations de HCHO (ppb) et humidité spécifique g/kg.

2.2.4 Les particules

La concentration en nombre de particules durant cette période n'est pas la même selon la taille des particules. Pour certaines gammes de particules, cette concentration varie en fonction de l'heure, notamment pour les particules fines. Le premier jour est marqué par des épisodes de deux pics de concentrations intérieures successives durant la première nuit (19/04/2010).

Avant l'apparition de ces pics de quelques minutes, un évènement très important s'est produit à l'extérieur : la concentration en nombre de particules à l'extérieur sur toutes les gammes fait brusquement changer l'allure générale des courbes de

concentrations. L'apparition des maxima pour les fines particules est globalement simultanée sur le pas de temps considéré, et des décalages de quelques minutes peuvent apparaître pour certains épisodes, toujours d'abord à l'extérieur ensuite à l'intérieur.

Le Tableau 2-1 suivant donne un aperçu sur l'intensité des pics pour les huit premières gammes de particules, calculée comme le rapport entre le maximum et la moyenne. Pour les particules intérieures de tailles 0.725 μm et 0.9 μm , le pic maximum représente 8 fois plus que la moyenne et il dure moins de 15 minutes. À l'extérieur, pour les particules de tailles supérieures à 0.725 μm , ce rapport atteint des valeurs supérieures à 40 ; notamment, pour les particules de taille 0.9 μm et 1.8 μm le maximum est 50 fois plus que la moyenne.

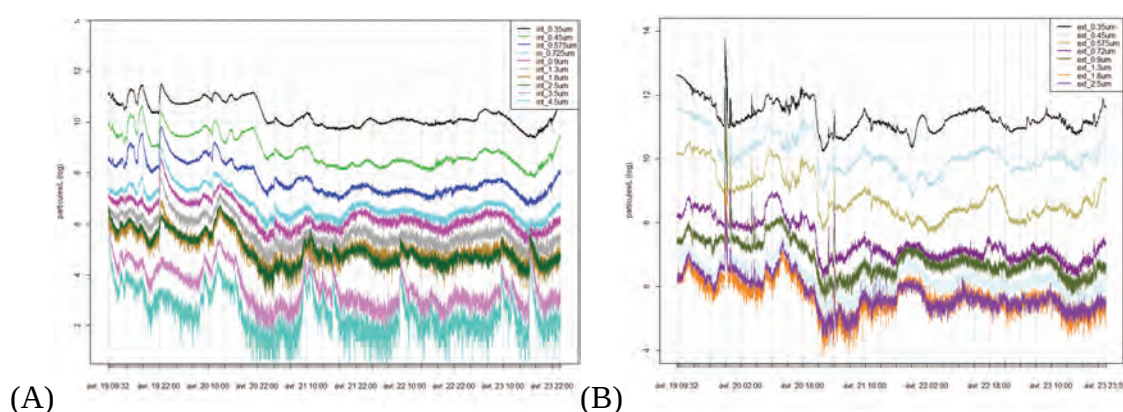
Tailles	0.35 μm	0.45 μm	0.57 μm	0.72 μm	0.9 μm	1.3 μm	1.8 μm	2.5 μm
Intérieur	3	4.7	6	8	8	7	5	4.7
Extérieur	10	10.5	21.5	41.7	51.7	41	51	44

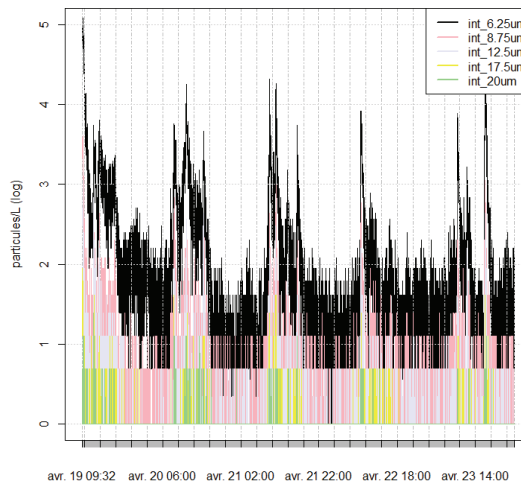
Tableau 2-2 Tableau récapitulatif du rapport entre le maximum et la moyenne pendant l'épisode pics.

La contribution majoritaire sur les variations de concentration intérieure pour les particules fines semble d'origine extérieure. En effet, le ratio moyen intérieur extérieur (I/E) est inférieur à 1 pour toutes les particules fines (0.35 μm à 2.5 μm).

Sur la Figure 2-5, on remarque que la variabilité temporelle pour ces particules présente deux régimes : au début (du 19/04 au 20/04 inclus), on constate une forte fluctuation et des concentrations importantes qui diminuent et varient peu par la suite. On peut remarquer un changement au niveau de la rose des vents, qui exhibe des différences d'un régime à l'autre, comme on peut le constater dans la Figure 2-6. Cela peut mettre en évidence une source importante de particules située sur le secteur Nord-Ouest et/ou au Sud.

En ce qui concerne les particules de taille supérieure à 4.5 μm , la variabilité temporelle présente un aspect différent de celui des particules plus fines, caractérisée par des discontinuités (Figure 2-5 bas). Cette discontinuité laisse présager les émissions des sources aléatoires.





(C)

Figure 2-5 (A): séries temporelles des concentrations en nombre de particules intérieures des dix premières gammes (en échelle logarithmique). (B) : séries temporelles des concentrations en nombre de particules extérieures des huit premières gammes (en échelle logarithmique). En bas : séries temporelles des concentrations en nombre de particules intérieures des 5 dernières fractions.

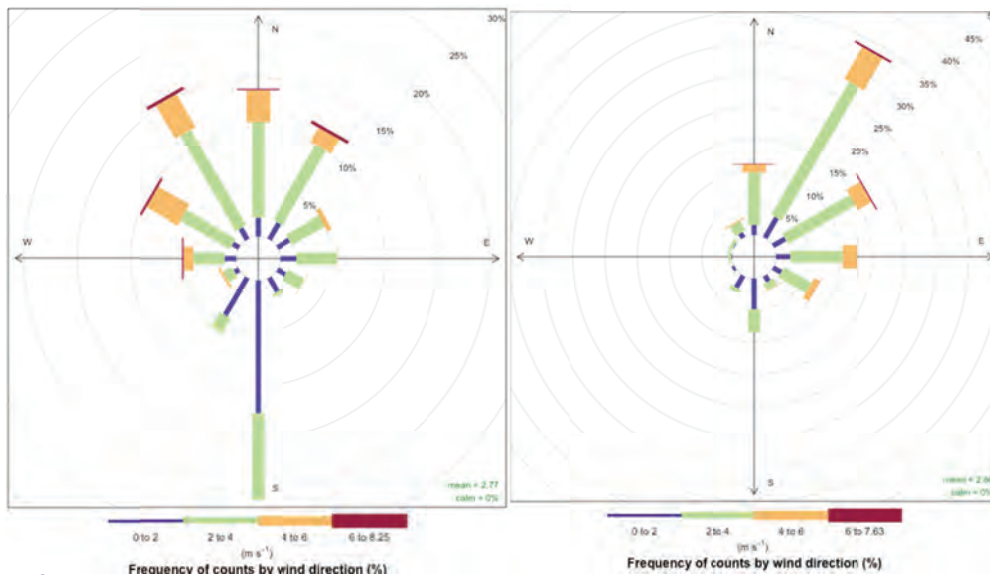


Figure 2-6 À gauche : rose des vents de la période (19 et 20 Avril) de fortes concentrations en particules fines (1^{er} régime des séries temporelles). À droite : rose des vents des concentrations de faible variabilité en particules fines (2^{ème} régime des séries temporelle)

Les résultats de l'évolution temporelle du ratio I/E sont illustrés dans la Figure 2-7. Une grande variabilité caractérise le ratio des particules moyennes (de 1.3 μm -2.5 μm). Des pics de concentrations apparaissent au cours du temps et leur fréquence s'accroît ponctuellement durant la première nuit (du 19/04/2010). La série temporelle du ratio synthétise de façon générale la consommation/production de la pollution particulaire sur certaines fractions. Par exemple, sur la période du 20-21/04/2010, on observe à l'intérieur une production significative de particules de tailles médiane 1.3 μm et 1.9 μm par rapport à l'ensemble de la période. Sur le graphique de la Figure 2-7, ces situations sont séparées par le seuil égal à 1.

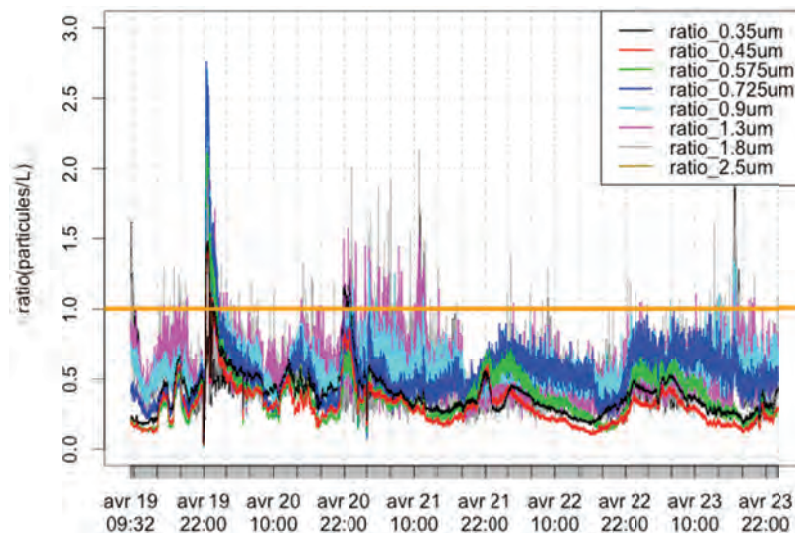


Figure 2-7: Évolution temporelle du ratio I/E (8 gammes de particules)

Sur l'ensemble des 8 gammes de tailles de particules comparées entre l'intérieur et l'extérieur, le ratio I/E reste inférieur à 1, voire inférieur à 0,5 pour les particules inférieures à $0,6\ \mu\text{m}$ de diamètre. Pour ces dernières, le ratio atteint un minimum au cours de la journée, qui est du au fait que le niveau extérieur augmente alors que le niveau intérieur reste stable (peu d'occupation ou d'activités à l'intérieur). Pour les particules comprises entre $0,8$ et $3\ \mu\text{m}$, le ratio I/E est relativement constant quelque soit l'heure de la journée. A noter pour les particules de $2,5\ \mu\text{m}$, un ratio I/E qui tend à légèrement augmenter durant la journée. Par ailleurs, les valeurs extrêmes se sont manifestées plutôt dans la nuit (22 h-00 h) pour les très fines particules ($0,35\ \mu\text{m}$ et $0,57\ \mu\text{m}$).

Tous ces résultats mettent en évidence la continuité de la contribution des sources extérieures sur les niveaux intérieurs dans la quasi-totalité de la période de mesure. Néanmoins, nous distinguons un phénomène atypique qui caractérise une source extérieure discrète dont nous allons estimer la contribution et caractériser la dynamique avec d'autres outils statistiques (§ 2.3). Ces observations rejoignent les conclusions de plusieurs travaux de recherches du rôle de la pollution atmosphérique extérieure sur les niveaux de pollution intérieure [2],[3].

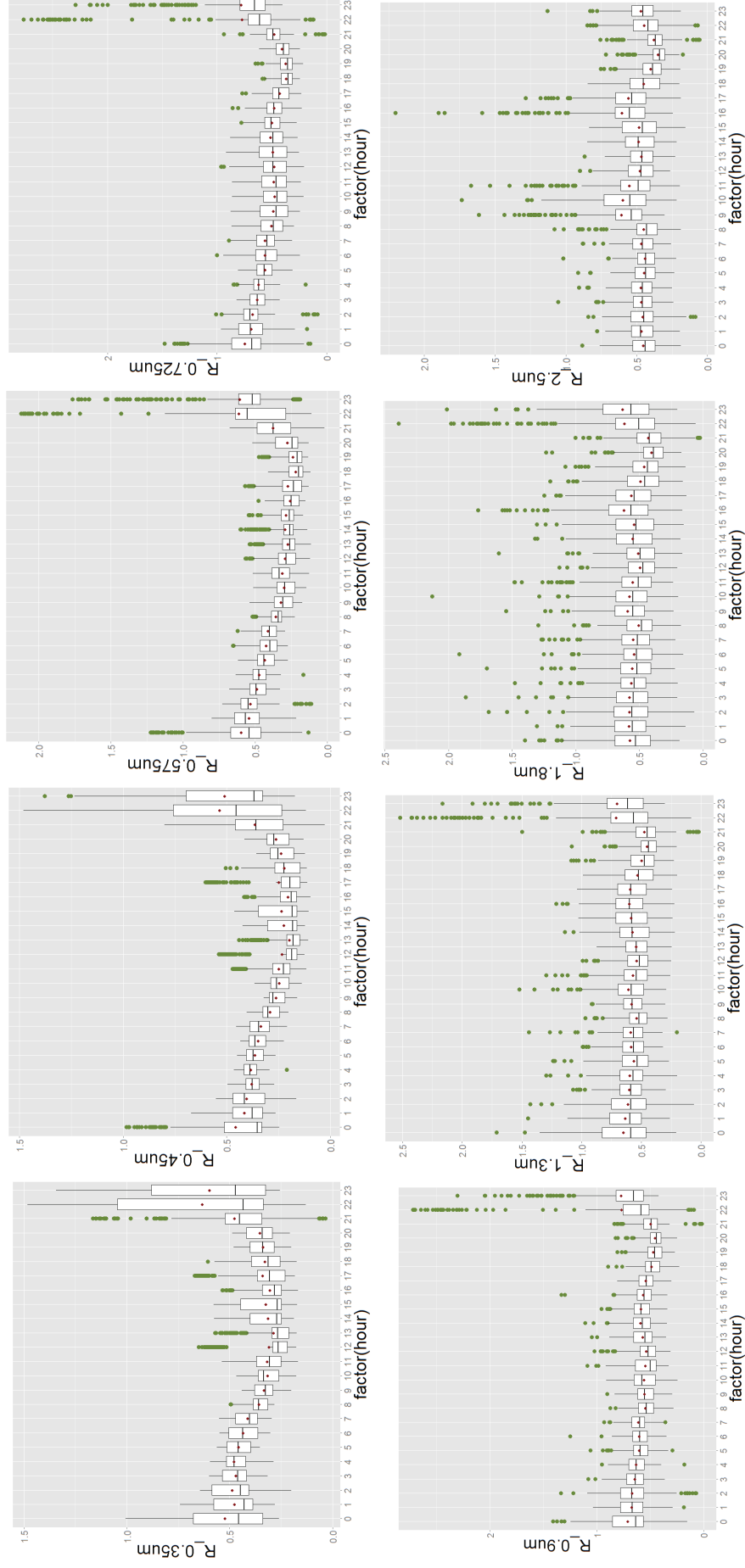


Figure 2-8 Profils diurnes des ratios Intérieur/Extérieur pour les différentes tailles de particules

2.2.5 Paramètres climatiques et de confort

Les paramètres climatiques sont de très grande importance dans l'interprétation des résultats sur le traçage et la contribution des sources extérieures. Dans cette étude, nous avons pris en compte notamment la vitesse et la direction du vent. La rose des vents présentée ici sont des données relatives aux mesures effectuées dans la station météo de la maison MARIA Figure 2.9.

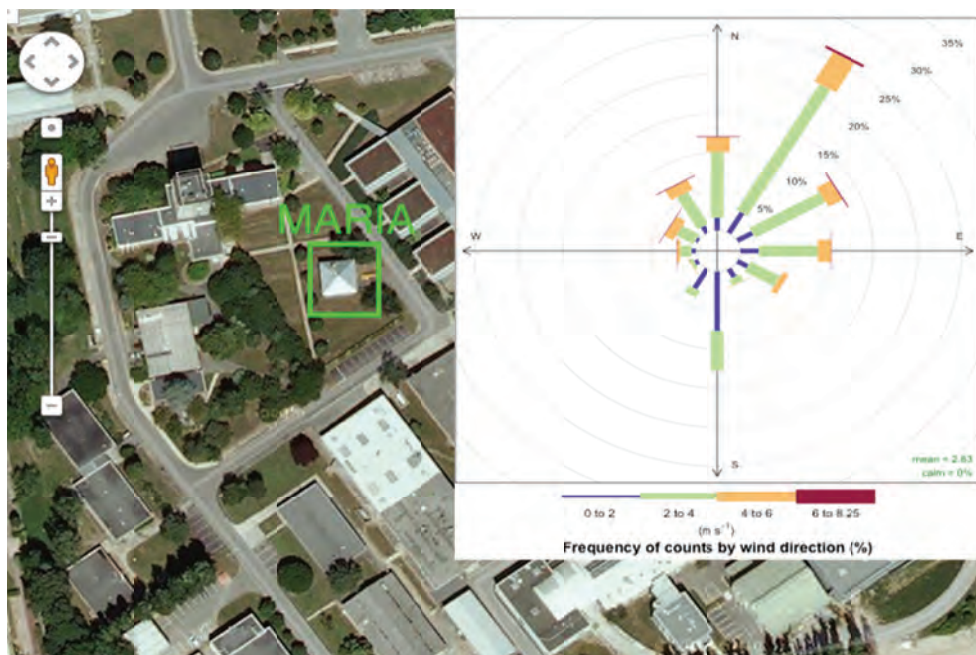


Figure 2-9 La rose des vents (vue aérienne) pendant la campagne des mesures

Sur la période du printemps pendant les quatre jours de mesures, la rose des vents indique que les vents proviennent principalement des directions Nord-Est et Est-Nord-Est et est caractérisée par une vitesse des vents de 2-6 m/s. Les données montrent aussi que des vitesses de 2-5 m/s provenant de la direction Nord et de la direction de Est (0°,90°). La vitesse du vent Sud (180°) est caractérisée par des vitesses faibles (0-3 ms^{-1}). Le secteur de vent 65° pour la direction Nord-Ouest favorise le transport des particules provenant des sources liées au trafic routier (routes D226 et D217).

Les données relatives aux autres paramètres étudiés concernent les niveaux du CO_2 , l'humidité et la température (intérieurs et extérieurs).

Les variations de teneurs en CO_2 intérieur vont de 360 ppm à 570 ppm avec des pics qui ne durent pas plus de 30 min. Sur cette période, l'occupation de la maison est caractérisée par une présence très faible.

Ayant une information supplémentaire sur le chauffage de la maison qui a été désactivé durant la campagne de mesures, on observe que l'écart de température est plus élevé à l'extérieur qu'à l'intérieur (31°C, 5.5°C). Cela est sans surprise, du fait que les variations extérieures suivent logiquement les conditions climatiques d'ensoleillement et que l'enveloppe de la maison possède une capacité à conserver la température.

2.2.6 Corrélation des polluants et classification

Les matrices de corrélations de Pearson et de Spearman nous donnent une première idée sur les relations linéaire (Pearson) ou type monotonie (Spearman) entre les rangs dans les séries temporelles (Figure 2-9). Pour l'interprétation des coefficients de Pearson, nous avons travaillé sur les variables brutes, même si pour certaines analyses, nous avons travaillé sur les logarithmes des variables.

De façon générale, le coefficient de Pearson obtenu indique que plus les particules sont fines, plus elles sont corrélées entre elles ; il varie pour ces tailles (PM intérieur) entre 0.80 et 0.95. Pour les particules intérieures de tailles allant de 0.35 μm à 0.57 μm , une corrélation forte a été détectée avec l'humidité relative intérieure, qui elle-même est fortement corrélée avec le CO_2 . Or, pour les niveaux extérieurs, aucune des corrélations significatives précédentes n'a été détectée entre ces paramètres, mises à part les particules de taille 0.45 μm avec la température extérieure et le CO_2 extérieur.

Pour le formaldéhyde : la corrélation de Pearson est positive pour la température intérieure (0.70), l'humidité spécifique intérieure (0.6) et négative pour l'humidité relative intérieure (-0.62). L'humidité spécifique quant à elle est fortement corrélée avec les particules de taille 0.3 μm intérieures et extérieures ; la plus forte corrélation a été enregistrée avec les particules intérieures de taille 2.5 μm et extérieures de taille 0.45 μm (0.69 et 0.67 respectivement).

En revanche, la corrélation de Spearman montre autre chose sur la forme et le type de relation qui peut exister entre toutes ces variables. En effet, bien que le coefficient de corrélation de Pearson ne dépasse que rarement 0.6 entre les particules intérieures et extérieures, la corrélation de Spearman est de l'ordre de 0.75 sur ces gammes de particules (< 2.5 μm).

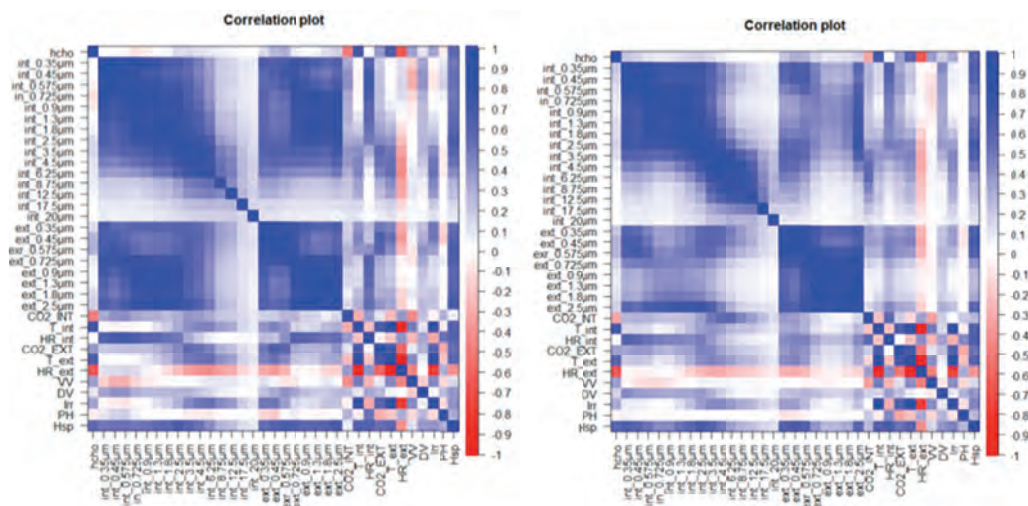


Figure 2-9: Corrélation entre les variables. À droite, la matrice de corrélation de Spearman. À gauche : la matrice de corrélation de Pearson. Corrélation positive forte (bleu foncé), corrélation négative (rouge).

Après cette analyse bidimensionnelle, on passe à une analyse multivariée.

Nous avons utilisé une analyse par classification ascendante hiérarchique (CAH) sur ces données afin de classer les polluants cibles (formaldéhyde, particules) ainsi que les différents paramètres climatiques dans des groupes homogènes par rapport à une métrique donnée.

Dans un premier temps, la variance des 35 séries temporelles a été neutralisée par la standardisation de la matrice. La deuxième étape consiste à prendre le logarithme de la translatée de la table de données ($X+1$, X étant la matrice de données), ceci évitera l'indétermination de logarithme 0 pour les grosses particules

En utilisant les données centrées réduites, les groupes obtenus peuvent être classés en 4 groupes principaux :

- **Classe 1** : regroupe les particules intérieures de tailles $0.35\ \mu\text{m}$ à $2.5\ \mu\text{m}$, l'humidité relative et l'humidité spécifique intérieur ;
- **Classe 2** : regroupe les particules de tailles supérieures à $4.5\ \mu\text{m}$ et toutes les particules extérieures.
- **Classe 3** : regroupe l'ensoleillement, le formaldéhyde, le CO_2 , la température intérieur et extérieure et la vitesse du vent ;
- **Classe 4** : regroupe le CO_2 , l'humidité relative extérieure et la pression atmosphérique.

2.3 Identification et contribution des sources

2.3.1 Formaldéhyde

Cette partie traite de l'estimation d'un mélange gaussien des données filtrées par une transformée de Fourier rapide (FFT). En effet, nous considérons que la densité de probabilité de la variabilité temporelle du HCHO totale est la résultante d'un mélange des processus sources. De plus, il est supposé que chaque source est caractérisée par une distribution de probabilité propre ; cette hypothèse est présentée dans [4] où Pearson fonde l'usage des mélanges : il justifie l'asymétrie d'un histogramme en supposant que l'échantillon qu'il étudie provient de deux populations gaussiennes distinctes et non d'une seule.

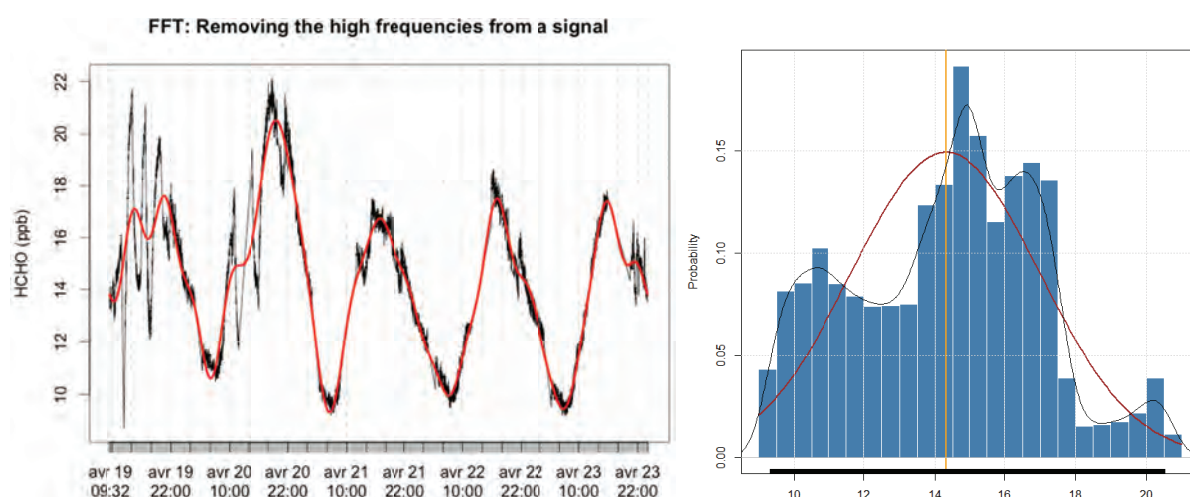


Figure 2-10 À gauche : Extraction des composantes de FFT appliquées aux variations de HCHO. À droite : histogramme de la composante FFT ainsi que l'estimation de la densité de probabilité de FFT de HCHO.

De fait de la forme quasi-sinusoïdale des concentrations observées (Figure 2-13), nous souhaitons à partir d'une transformée de Fourier rapide (FFT) garder que la structure « déterministe » de la dynamique des fluctuations, ensuite estimer les différentes densités de probabilités qui la constituent et qui sont mélangés. La méthode statistique utilisée pour ajuster le modèle aux données est l'estimation par

maximum de vraisemblance. Les algorithmes de calcul sont basés sur une méthode de type Newton et l'algorithme est celui de "Expectation Maximization" (EM). Dans un premier temps, un ajustement a été réalisé en faisant une hypothèse de normalité des distributions de la composante de Fourier. Un autre ajustement a été réalisé avec le logarithme de la composante de Fourier.

L'estimation du nombre de modes est basée sur le critère de BIC (Bayesian Information Criterion) [5]. Les différentes simulations pour l'optimisation du critère BIC ont été fait sans contrainte sur les proportions, moyennes et écarts types de différentes composantes que peuvent constituer la densité de probabilité totale.

La (Figure 2-11, à gauche) montre l'ajustement de mélanges de lois normales appliqué au jeu de données de fréquence de la composante de Fourier du HCHO. Les lois de distribution estimées sont différentes selon la paramétrisation utilisée (ici nous ne présentons pas la complexité des algorithmes et le temps de calcul). La (Figure 2-11, à droite) représente les deux composantes estimées dans le cas d'une transformation logarithmique de la composante de Fourier. La (Figure 2-11, à gauche) représente cinq densités de probabilité estimées pour la composante de Fourier.

L'ajustement de la distribution polymodale donne des distributions relativement cohérentes du point de vue qualité de l'air intérieur. En effet, les composantes 1 et 4 sont manifestement différentes. On supposera donc pour classier les sources et leurs contributions qu'elles sont générées par de mélanges provenant d'émissions distinctes. La contribution de chaque composante est donnée dans le Tableau 2-3.

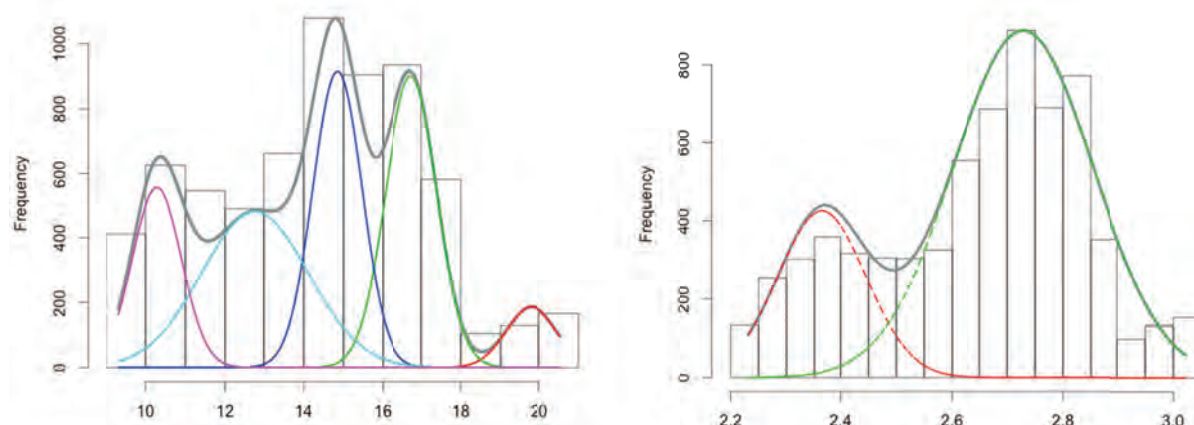


Figure 2-11 À gauche : Histogramme des fréquences de la composante déterministe de la variabilité de FFT du HCHO et les lois de distribution des fréquences estimées. À droite : L'estimation après transformation en Log.

	Composante par classe	Moyenne	variance	Proportion %
FFT(HCHO)	Cluster 1 (rouge)	19.79	0.41	5,3
	Cluster 2 (vert)	16.73	0.43	26
	Cluster 3 (bleu)	14.85	0.38	25
	Cluster 4 (turquoise)	12.76	1.82	28
	Cluster 5 (rose)	10.28	0.38	15
Log(FFT (HCHO))	Cluster 1 (vert)	2,72	0.015	76,48
	Cluster 2 (rouge)	2.36	0.0065	23,53

Tableau 2-3 Estimation des composantes de mélange pour la composante de Fourier.

Le Tableau 2-3 synthétise l'estimation de la moyenne, variance et contribution de différentes composantes

2.3.2 Particules

Dans cette partie, une étude sur la caractérisation temporelle des fluctuations de particules générées par différentes sources a été entreprise. Plus précisément, il s'agit de caractériser la structure temporelle des sources potentielles de particules. Pour cela, nous avons d'abord entrepris une analyse en composante indépendante pour les ratios afin d'en dégager l'évolution temporelle de la variabilité de la source extérieur et sa contribution. Ensuite, la même analyse a été effectuée pour les particules intérieures. Notons que les modèles d'identification « en aveugle » ont été essentiellement développés dans le domaine de traitement du signal ; néanmoins, plusieurs méthodes (mais pas l'ACI) ont été appliquées dans l'air extérieur : par exemple l'analyse en composantes principales (ACP) ou la Factorisation Matricielle Positive (PMF)...

Comme pour l'identification par séparation de mélange densités gaussiennes appliquée au formaldéhyde, nous ne présenterons pas les développements mathématiques sous-jacents. Nous ferons aussi abstraction des propriétés statistiques de quelques concepts qui fondent certains algorithmes de l'ACI, notamment la néguentropie. Pour une présentation globale de ces méthodes, nous renvoyons le lecteur aux références [6]

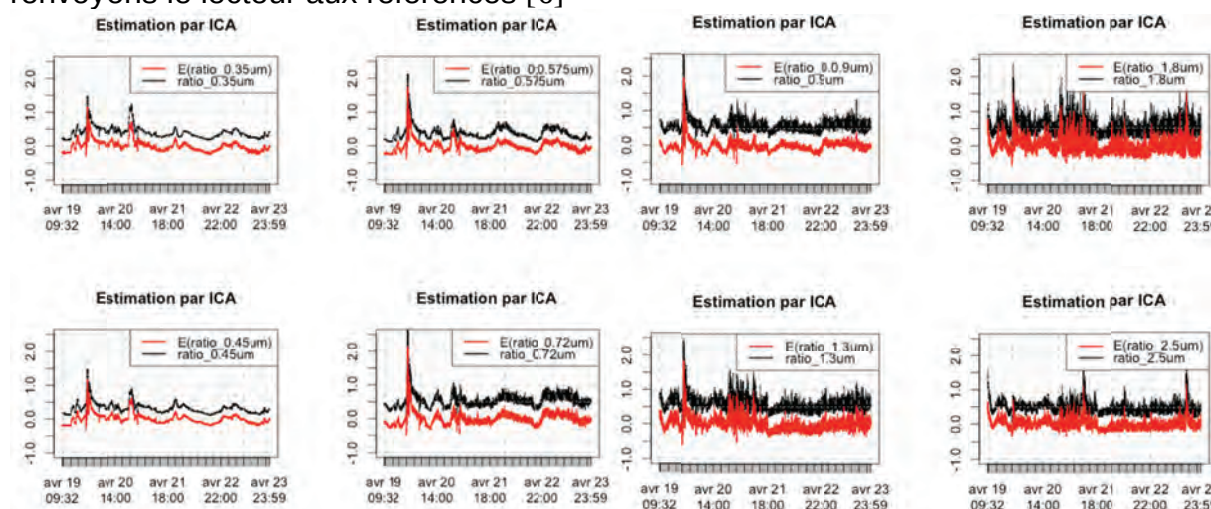


Figure 2-12 Estimation du ratio I/E de particules par une analyse en composante indépendante

L'algorithme utilisé dans cette étude pour la déconvolution aveugle par l'ACI nécessite de spécifier le nombre de sources. Dans le cas de traitement des ratios, nous avons délibérément choisi deux composantes pour huit séries chronologiques. La Figure 2-12 représente les huit ratios utilisés ainsi que l'estimation par décomposition en composantes indépendantes. Les résultats présentent une légère sous-estimation des données originales mais suivent la même dynamique

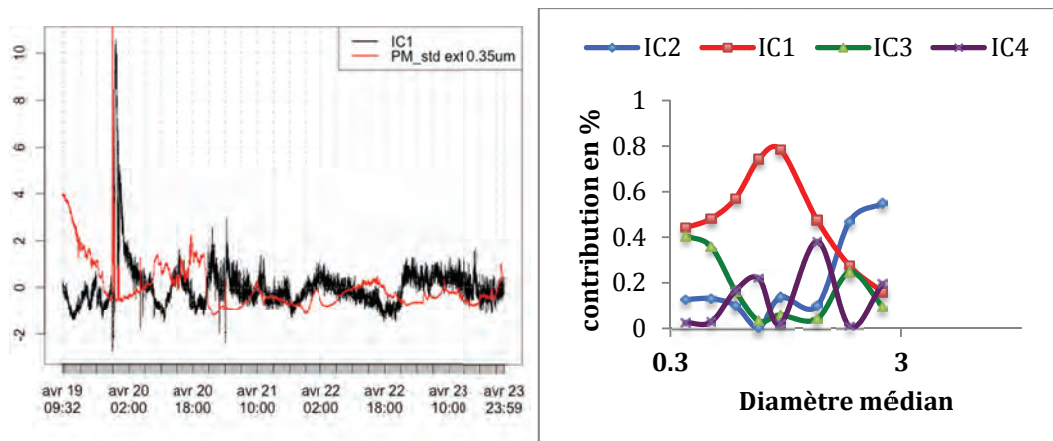


Figure 2-13 À gauche : l'évolution temporelle de la première composante indépendante obtenue sur les données ratio des particules, ainsi que l'évolution de la série temporelle de la concentration en nombre de particules de taille 0.35 μm. À droite la contribution des 4 sources estimées sur les 8 gammes de ratios.

Dans la Figure 2-13 à droite, on présente la contribution des sources estimée par l'ACI pour chaque composante indépendante sur chaque fraction du ratio de particules. De façon générale, la contribution des sources est distributive avec une source majoritaire (composante 1) qui représente les émissions provenant de l'extérieur. Elle représente plus de 40 % pour les ratios de gammes de 0.35 μm, 0.45 μm et 1.3 μm ; plus de 50 % pour la taille 0.57 μm et avoisinent les 80% pour les ratios de tailles 0.72 μm et 0.9 μm. Elle représente, en plus, une source moins contributive pour les grosses particules (> 2.5 μm). Dans la Figure 2-13 à gauche, on présente la série temporelle de la première composante indépendante et de la série en nombre de particule de taille 0.35 μm. La composante 2 représente sur cette gamme de particules plus de 50% de l'ensemble des autres contributions de sources. De plus, la contribution de la source 2 est relativement faible pour les plus fines particules, mais avoisine les 50 % des émissions totaux pour les particules de tailles 1.8 μm et plus de 50 % pour les gammes 2.5 μm. La composante 3 part sur la même importance que la composante 1 dans la définition du ratio pour les particules de tailles 0.3 μm, 1.8 μm et 2.5 μm ; mais dans un sens inverse pour les tailles entre 0.57 μm à 1.3 μm.

En résumé, la contribution des sources sur le rapport intérieur et extérieur définit l'importance des processus émissions, dépôt et remise en suspension.

La deuxième étape consiste à appliquer l'ACI sur l'analyse en composantes indépendantes sur toutes les particules intérieures et extérieures. Le choix du nombre de composantes retenues est basé sur la recherche du premier minimum local de l'erreur quadratique moyenne. On remarque qu'en se basant sur ce critère, le nombre de composantes à retenir pour chaque fraction varie en fonction de la taille médiane des particules. Nous retiendrons au finale que 4 composantes pour l'ACI afin d'éviter le sur-apprentissage de la décomposition.

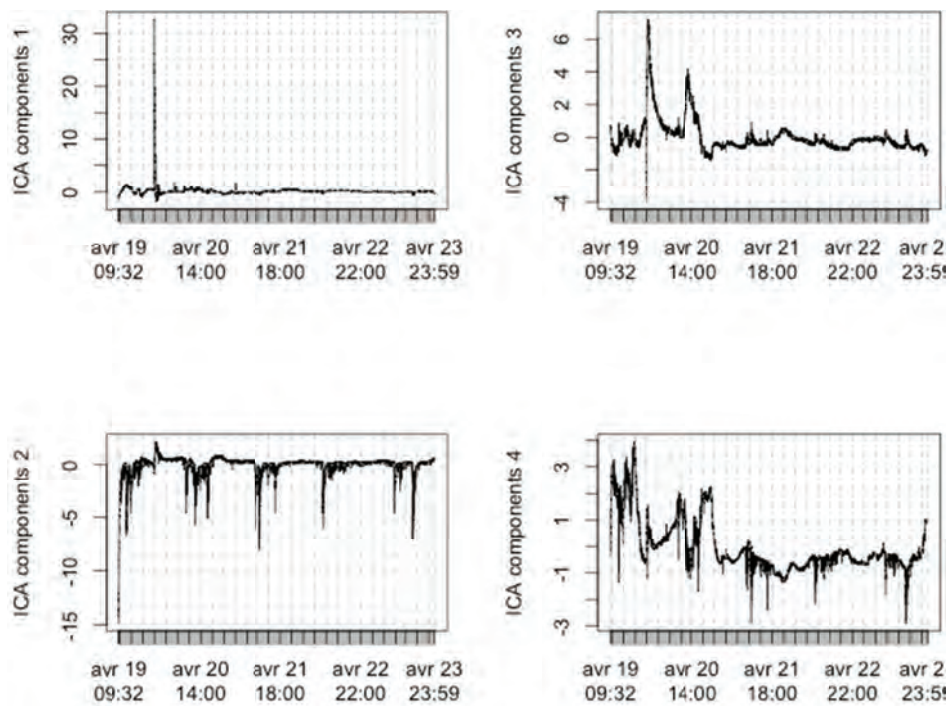


Figure 2-14 Profils temporels des quatre premières composantes

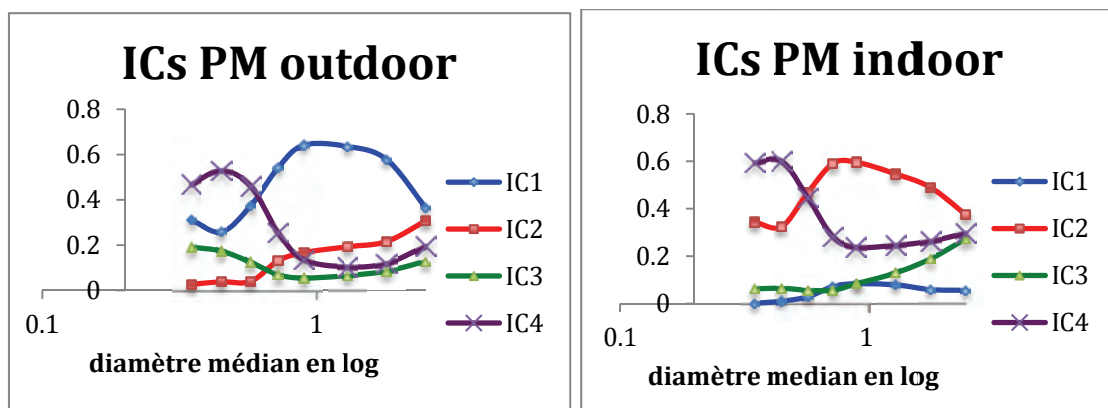


Figure 2-15 Contribution des sources sur les huit premières gammes de particules.

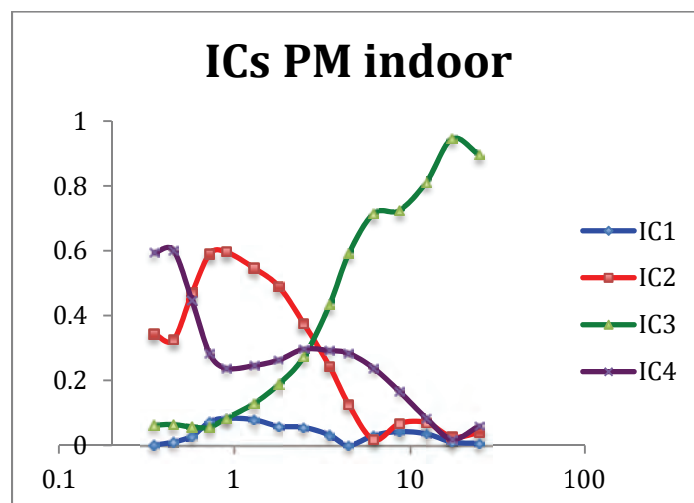


Figure 2-16 Contribution des sources sur les 15 gammes de particules intérieures

La Figure 2-14 montre les fluctuations temporelles des quatre composantes indépendantes obtenues sur les 24 fractions de particules. La deuxième composante obtenue est inversée en signe, mais ceci n'est qu'un artefact, car les solutions sont déterminées à un facteur près, donc la multiplication par (-1) est envisageable. La variabilité de la première composante indépendante ressemble beaucoup à celle observée pour les particules extérieures, notamment pour les fines ($0.35\ \mu\text{m}$ à $0.7\ \mu\text{m}$), avec un pic très élevé durant la fin du premier jour ; la composante 1 laisse penser qu'elle est représentative l'ensemble des contributions extérieures.

De plus, les fluctuations de la composante 4 possèdent le même type de variabilité que les fines particules intérieures.

Dans la Figure 2-15, on présente la contribution des sources estimée par l'ACI pour chaque composante indépendante sur les 8 fractions de particules intérieures et extérieures. Les quatre composantes contribuent de façon distributive en fonction de la taille des particules. Pour les fractions les plus fines ($0.35\ \mu\text{m}$, $0.45\ \mu\text{m}$), la contribution de composante 4 (C4) est d'environ 50% dans l'air intérieur et plus de 60% dans l'air extérieur. La contribution de la (C4) diminue sur les fractions supérieures à $0.57\ \mu\text{m}$. et elle devienne presque nulle pour les grosse particules (supérieur à $15\ \mu\text{m}$, voir la Figure 2-1619). Bien que la composante 1 ne contribue pas dans l'air intérieur, elle est majoritaire sur certaines fractions dans l'air extérieur. Notons aussi l'aspect qualitatif entre la distribution des contributions de la composante 2 pour l'extérieur et de la composante 1 pour l'intérieur.

La contribution de la composante 3 caractérise l'importance des processus d'émission et de remise en suspension dans l'air intérieur. En effet, l'évolution de sa contribution suit un schéma type d'une source indépendante des teneurs extérieur, elle est majoritaire pour les grosses particules et contribue moins aux teneurs des particules fines intérieures. Néanmoins, on observe une contribution de 20 % de la C3 sur les fractions fines extérieures.

3 Conclusion

Face à la problématique de l'identification et de la variabilité des sources et de leur contribution, deux méthodes ont été utilisées : la séparation par un modèle de mélange de densités gaussiennes appliquée aux fluctuations du HCHO et l'analyse en composantes indépendantes appliquée aux fluctuations intérieures et extérieures des particules.

Cette étude a permis de mettre en évidence la contribution des sources intérieures et extérieures dans un contexte de la qualité de l'air intérieur d'une maison expérimentale. Pour le formaldéhyde, il est très difficile de caractériser les émissions chaque source, la méthode nous permet juste de quantifier les fréquences d'occurrence de chaque classe d'émissions.

En définitive, avec un nombre de composantes assez réduit (3-4) par rapport au nombre de variables (24), on arrive à reconstituer de manière très satisfaisante les variables initiales. Ce résultat est une étape importante pour la prévision des concentrations de polluants qui pourra se faire par les réseaux de neurones à partir des composantes uniquement, tout en évitant la redondance entre les entrées.

Les différents profils temporels des composantes obtenues sont interprétés en termes de sources de pollution. Dans ce contexte on a pu remarquer que le profil temporel de certains polluants est corrélé avec les composantes indépendantes obtenues. Ce travail sera utilisé ensuite comme base pour la prévision des polluants dans l'air intérieur.

Bibliographie

- [1] Koffi J., Ribéron J., Allard F., Experimental study of pollutant spread within dwellings. Proceedings of Roomvent 2007, 10th International Conference on Air Distribution in Rooms, Helsinki 13-15 June 2007.
- [2] Morawska L, Thomas S, Gilbert D, Greenaway C, Rijnders EA. A study of the horizontal and vertical profile of submicrometer particles in relation to a busy road. Atmospheric Environment 1999 ; 33 (8) :1261-74.
- [3] Zhu YF, Hinds WC, Krudysz M, et al. Penetration of freeway ultrafine particles into indoor environments. Journal of Aerosol Science 2005 ; 36 (3) : 303-22.
- [4] Pearson. K. Contributions to the theory of mathematical evolution. Philosophical Transactions of the royal Society of London A, 185 :71–110, 1894.
- [5] Schwarz, G. Estimating the dimension of a model. The Annals of Statistics, 6 :461–464, 1978.
- [6] Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis". IEEE Transactions on Neural Networks , 1999 10 (3): 626–634. doi:10.1109/72.761722